(19)日本国特許庁（ＪＰ）　(12) **公開特許公報**（Ａ）　(11)特許出願公開番号
特開2000－56797
（P2000－56797A）
(43)公開日　平成12年２月25日(2000.2.25)

(51)Int.Cl.'		識別記号	Ｆ Ｉ			ﾃｰﾏｺｰﾄﾞ(参考)
Ｇ１０Ｌ	15/28		Ｇ１０Ｌ	3/00	５７１Ｋ	５Ｄ０１５
	13/00				Ｒ	５Ｄ０４５
	11/02				５１３Ｚ	９Ａ００１
Ｇ０６Ｆ	3/16	３２０	Ｇ０６Ｆ	3/16	３２０Ｈ	
Ｇ１０Ｌ	11/00		Ｇ１０Ｌ	9/00	Ａ	

審査請求　未請求　請求項の数３　ＯＬ　（全　７　頁）　最終頁に続く

(21)出願番号　特願平10－224927

(22)出願日　平成10年８月７日(1998.8.7)

(71)出願人　000000376
オリンパス光学工業株式会社
東京都渋谷区幡ヶ谷２丁目43番２号

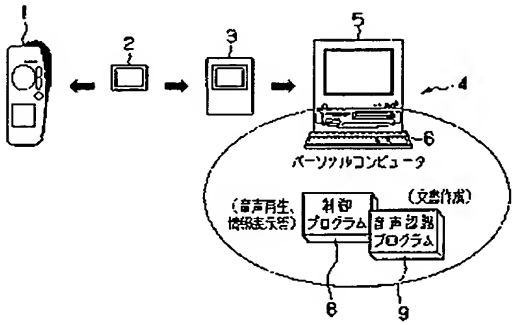
(72)発明者　▲高▼橋　秀享
東京都渋谷区幡ヶ谷２丁目43番２号　オリンパス光学工業株式会社内

(74)代理人　100076233
弁理士　伊藤　進

Ｆターム(参考)　5D015 DD00 LL09
5D045 AB02 DB01
9A001 BZ03 HH16 HH17 JJ28 JJ29

(54)【発明の名称】　音声処理装置

(57)【要約】
【課題】よりユーザフレンドリーな音声処理装置を提供する。
【解決手段】ミニチュアカード２に記録された、ディジタルレコーダ１で記録した音声データをパーソナルコンピュータ４で読み込み、このパーソナルコンピュータ４で、読み出した音声データの信号対雑音比を演算し、信号対雑音比の値が適正かどうかを判断し、信号対雑音比の値が適正である場合に音声データを音声認識プログラム９により音声認識する。



1

【特許請求の範囲】
【請求項１】　音声データが記録された記録媒体から音声データを読み出す音声データ読出手段と、
上記音声データ読出手段で読み出した音声データの信号対雑音比を演算する信号対雑音比演算手段と、
上記信号対雑音比演算手段の出力値が適正であるか否かを判断する信号対雑音比判断手段と、
上記信号対雑音比判断手段で上記出力値が適正であると判断した場合に、上記音声データを音声認識処理する音声認識処理手段と、
を具備することを特徴とする音声処理装置。
【請求項２】上記信号対雑音比判断手段で上記出力値が適正でないと判断した場合に、該適正でない旨を表示する警告手段をさらに具備することを特徴とする請求項１に記載の音声処理装置。
【請求項３】上記信号対雑音比判断手段で上記出力値が適正でないと判断した場合に、上記音声データに音声認識処理を施すか否かを選択する選択手段をさらに具備することを特徴とする請求項１記載の音声処理装置。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】本発明は、音声処理装置、詳しくは、記録媒体から読み出した音声データに所定の処理を施す音声処理装置に関する。
【０００２】
【従来の技術】いわゆる音声ワードプロセッサ、あるいは、口述することにより音声データを入力すると該音声データに基づいて自動的に文言を作成しそれを画面等に表示する、いわゆるディクテーションシステムの実現は、従来からの音声認識システム開発における一つの目標であり、現在、活発に研究が進められている。
【０００３】こうした近年の音声認識技術および計算機技術の進歩に伴って、パーソナルコンピュータにマイクロフォンを接続し、このマイクロフォンを用いて入力した音声を該パーソナルコンピュータ上で文言化して画面に表示させる装置が開発されていて、一般に市販されている。
【０００４】一方、従来より、文言を作成するにあたり作成したい文書の内容をいったんテープレコーダ等の録音装置に口述録音して、後で秘書やタイピスト等がその口述内容を再生しながらタイプライタやワードプロセッサ等の文言作成装置により文言化する、といった形態をとることが、テープレコーダ等の録音装置の有効な利用形態の１つとして一般化している。
【０００５】このような口述録音においては以前から、録音内容を自動的に文言に変換する技術の実現が強く望まれている。
【０００６】また、近年のコンピュータ技術、ディジタル信号処理技術等の発展により、録音内容をディジタルデータ化して、フラッシュメモリ等の書込み、消去が可

2

能な記録媒体に記録する、いわゆるディジタルレコーダが開発されるようになり、さらに、そのディジタル化された録音内容をパーソナルコンピュータに転送して、該パーソナルコンピュータにおいて録音内容を再生することが可能となっている。
【０００７】本出願人は、このようなディジタルレコーダから転送された録音データを、パーソナルコンピュータ上で簡単において簡単な操作で扱うことを可能とする音声データの処理制御装置を開発しており、特願平９－１４９７２８号において提案している。
【０００８】さらに本出願人は、ディジタル記録された音声を、上記音声データの処理制御装置から音声認識装置に渡して音声認識させ、文言として画面に表示させるディクテーションシステムを開発しており、特願平９－１４９７２９号において提案している。
【０００９】このようなディクテーションシステムによれば、コンピュータの前に座って直接音声入力をする必要がなく、いったんディジタルレコーダに録音して、後でコンピュータにその録音データを転送して文書を作成させることが可能となる。
【００１０】
【発明が解決しようとする課題】ところで、上述したようなディクテーションシステムにおける音声認識処理には、不特定話者向け大語彙連続音声認識技術が必要となる。しかしながら、現在の不特定話者向け大語彙連続音声認識技術においては、誤認識のない、完璧な認識結果を得ることは大変難しく、特に、認識対象の音声中に背景雑音が混入してしまうと認識性能が劣化するという問題がある。従来、このような問題を解決するために様々な提案がなされているのはよく知られているところである。しかしながら、限られた装置でそれを解決することは困難である。
【００１１】このような現状の中で、上述したようなパーソナルコンピュータにマイクロフォンを接続し、このマイクロフォンを用いて入力した音声を該パーソナルコンピュータ上で文言化して画面に表示させる装置を使用する場合にあっては、その場で画面に表示される音声認識結果を確認して、使用者の判断で、誤認識が多ければ、再び音声入力をやり直すといった対応をとることも可能である。
【００１２】一方、上述したような、ディジタル記録された音声データを処理制御装置から音声認識装置に対して音声認識をさせ、認識した結果を文書として画面に表示させるディクテーションシステムにおいては、すでに記録された音声データが音声認識装置への入力となる。
【００１３】そのために、大きな背景雑音が混入して記録された音声データに対して音声認識処理を行なうと、誤認識が多く、後で修正することさえも困難な認識結果が表示されてしまい、使用者の判断で再び音声認識処理を実行し直しても認識結果が改善される見込みがないと

いう問題がある。

【００１４】上述したようなディクテーションシステムが本来の目的とするところは、記録された音声データの内容をより速く、より簡単に文書化すること、すなわち文書作成支援を行うことにある。記録された音声データに対する音声認識処理の結果、誤認識部分がわずかであれば、その誤認識部分をキーボード、マウス等を用いて修正するだけで済むので、その目的を達することができる。

【００１５】しかし、ある程度以上に誤認識部分が増えると修正さえも困難となり、結局、初めからタイプし直したほうが速く文書を作成できるということになってしまう。このようなことは実際に処理して結果を出してみないと分からない、というのでは使用者に大きな不便を負わせることになってしまう。

【００１６】本発明はかかる問題点に鑑みてなされたものであり、よりユーザフレンドリーな音声処理装置を提供することを目的としている。

【００１７】

【課題を解決するための手段】上記の目的を達成するために本発明の第１の音声処理装置は、音声データが記録された記録媒体から音声データを読み出す音声データ読出手段と、上記音声データ読出手段で読み出した音声データの信号対雑音比を演算する信号対雑音比演算手段と、上記信号対雑音比演算手段の出力値が適正であるか否かを判断する信号対雑音比判断手段と、上記信号対雑音比判断手段で上記出力値が適正であると判断した場合に、上記音声データを音声認識処理する音声認識処理手段と、を具備する。

【００１８】上記の目的を達成するために本発明の第２の音声処理装置は、上記第１の音声処理装置において、上記信号対雑音比判断手段で上記出力値が適正でないと判断した場合に、該適正でない旨を表示する警告手段をさらに具備する。

【００１９】上記の目的を達成するために本発明の第３の音声処理装置は、上記第１の音声処理装置において、上記信号対雑音比判断手段で上記出力値が適正でないと判断した場合に、上記音声データに音声認識処理を施すか否かを選択する選択手段をさらに具備する。

【００２０】

【発明の実施の形態】以下、図面を参照して本発明の実施の形態を説明する。

【００２１】図１乃至図６は、本発明の一実施形態であるディクテーションシステムにかかり、図１は、該ディクテーションシステムの概念的な全体構成を示した図である。

【００２２】このディクテーションシステムは、図１に示すように、音声を電気信号に変換して音声データ化するディジタルレコーダ１と、このディジタルレコーダ１に着脱可能に装着して用いられるものであって上記音声データを記録する記録媒体たるミニチュアカード２と、このミニチュアカード２を後述するＰＣカードスロット４０（図２参照）に挿入して接続可能とするためのＰＣカードアダプタ３と、出力手段たるディスプレイ５やキーボード６，マウス７等を備え、上記ＰＣカードスロット４０を介して上記ミニチュアカード２から得た音声データに、制御プログラム８や音声認識プログラム９による処理を施す音声処理装置としてのパーソナルコンピュータ４とを有して構成されている。

【００２３】なお、上記ディスプレイ５は、後述する信号対雑音比が適正でない旨を表示する警告手段としての役目を果たす。

【００２４】図２は、上記パーソナルコンピュータ４の電気的な構成を示すブロック図である。

【００２５】このパーソナルコンピュータ４は、上記制御プログラム８に従って音声再生や情報表示等を行い、また上記音声認識プログラム９に従って文書作成等を行うとともに、その他の各種のプログラムに応じて様々な処理を行うものであって、当該パーソナルコンピュータ４全体の制御を司ると共に、音声データ読出手段、信号対雑音比演算手段、信号対雑音比判断手段、音声認識処理手段等の役目を果たすＣＰＵ３１と、このＣＰＵ３１の作業領域となる記録媒体たるメインメモリ３２と、例えばハードディスクやフロッピーディスク等でなり上記制御プログラム８や音声認識プログラム９が記録されている記録媒体たる内部記録媒体３３と、各種の外部機器に接続するための外部ポート３４と、上記ディスプレイ５を接続するインターフェース（以下、ＩＦと略す）３５と、上記キーボード６やマウス７を接続するＩＦ３６と、音声データに基づいて音声を発するスピーカ３８と、このスピーカ３８を接続するＩＦ３７と、上記ＰＣカードアダプタ３に装着されたミニチュアカード２が挿入されるＰＣカードスロット４０と、このＰＣカードスロット４０を接続するためのＩＦ３９と、を有して構成されている。

【００２６】また、上記ＣＰＵ３１、メインメモリ３２、内部記録媒体３３、外部ポート３４、ＩＦ３５，３６、３７，３９は、バスを介して互いに接続されている。

【００２７】なお、音声データは、上記ＰＣカードスロット４０を介してミニチュアカード２から直接読み込むようにしても良いが、一旦、上記内部記録媒体３３に記録して、この内部記録媒体３３から読み出すようにしても良いし、あるいは、ディジタルレコーダ１から通信手段等を介して直接読み込むようにしても構わない。

【００２８】また、上記スピーカ３８は、信号対雑音比が適正でない旨を表示（発音）する警告手段としての役目を果たす。

【００２９】次に、本実施形態のディクテーションシステムにおける音声認識処理を、図３，図４を参照して説

明する。

【００３０】図３は、本実施形態のディクテーションシステムにおいて音声メモリから音声データを読み出して音声認識するときの全体の流れを示す概念図であり、図４は同ディクテーションシステムにおける音声認識の処理を示すフローチャートである。

【００３１】図４に示すように、音声処理を開始すると、上記ミニチュアカード２または上記内部記録媒体３３等、音声データが記録された記録媒体としての音声メモリ１１（音声データが記録された記録媒体）からファイル単位で記録されている音声データを読み込み（ステップＳ１）、復号化処理１２を行う（ステップＳ２）。

【００３２】この復号化処理１２の結果はＳＮ比（信号対雑音比）計算処理１３に送られ、例えば後述する手法でＳＮ比の計算を行う（ステップＳ３）。

【００３３】このＳＮ比の計算値（Ｓ／Ｎ）は、判定処理１４により所定値ｍと比較される（ステップＳ４）。このステップＳ４において、Ｓ／Ｎ＞ｍであれば音声データが音声認識処理１５に送られて、音声認識が行われる（ステップＳ５）。そしてこの音声認識の結果をディスプレイ５等の画面に表示する（ステップＳ６）。

【００３４】上記ステップＳ４で、Ｓ／Ｎ＞ｍでなければ、このまま音声認識を続行してもよい認識結果が得られないだろうという旨の警告をディスプレイ５に表示する（ステップＳ７）。

【００３５】なお、この警告表示はディスプレイ５に表示させるに限らず、例えばスピーカ３８より音声等で警告しても良い。

【００３６】上記の警告の後、音声認識処理をこのまま実行するか、処理の実行を止めるか、使用者に選択を促す（ステップＳ８）。なお、この選択は、キーボード６上のキー操作によっても良く、また、マウス７等によって操作されても良い。

【００３７】上記使用者による選択の結果、ｙｅｓであればステップＳ５に行って、音声認識処理を実行する。一方、使用者による選択の結果、ｎｏであれば、処理を終了する。

【００３８】次に、上記ステップＳ３におけるＳＮ比計算処理の処理内容について、図５、図６に示すフローチャートを参照して説明する。

【００３９】まず、音声データの雑音レベルの計算手法を図５に示すフローチャートを参照して説明する。

【００４０】この処理が始まると、まず、フレーム番号のカウント値を示す変数ｆを０に初期化しておく（ステップＳ１１）。

【００４１】次に、変数ｆをインクリメントした後に（ステップＳ１２）、図示の数式によりフレームエネルギーｅ（ｆ）を計算する（ステップＳ１３）。なお、数式中、ｓ（ｉ）は１フレーム中の（ｉ）番目のサンプルにおける入力信号、Ｎは１フレームを構成するサンプル数を示している。

【００４２】次に、変数ｆの値が１であるか否か、すなわち、初期のフレームであるか否かを判定し（ステップＳ１４）、ｆが１である場合には、最小フレームエネルギーを示す変数ｍｉｎの値をｅ（１）にセットする（ステップＳ１６）。

【００４３】また、上記ステップＳ１４においてｆが１でない場合には、フレームエネルギーｅ（ｆ）が変数ｍｉｎより小さいか否かを判定し（ステップＳ１５）、小さい場合には変数ｍｉｎにフレームエネルギーｅ（ｆ）をセットし（ステップＳ１７）、一方、小さくない場合にはそのまま何もせずに次のステップＳ１８に行く。

【００４４】そして、ファイルが終端に達したか否かを判定し（ステップＳ１８）、まだ終端でない場合には上記ステップＳ１２に戻って上述した処理を繰り返す。

【００４５】また、このステップＳ１８においてファイルの終端に達したと判断された場合は、しきい値ｎｏｉｓｅｌｅｖとして、上記変数ｍｉｎに所定の値α（例えば１．８）を積算した値をセットして（ステップＳ１９）、この処理を抜ける。

【００４６】このような雑音レベルの検出方法は、すでに音声データが記録されていることを有効に利用したものであり、精度のよいＳＮ比の計算に資することが可能となる。

【００４７】なお、上述では、読み込んだ全区間（つまり、音声ファイルを構成する全フレーム）の最小値を求めているが、本発明はこれに限定されるものではなく、全ての区間の最小値でなくても、ある程度の長さの区間であれば良い。

【００４８】次に、ＳＮ比を計算する処理の内容を図６に示すフローチャートを参照して説明する。

【００４９】この処理が始まると、フレーム番号のカウント値を示す変数ｆ、各フレームの信号対雑音比（ＳＮ比）の加算値を示す変数Sum、加算回数を示す変数Cntを、各々"０"に初期化しておく（ステップＳ２１）。

【００５０】次に、変数ｆをインクリメントして（ステップＳ２２）、上述した図５において計算したフレームエネルギーe(f)が、雑音レベルnoiselevより大きいか否かを判定する(ステップＳ２３)。ここでe(f)がnoiselevよりも大きい場合には、当該フレームの信号対雑音比の計算値を変数Sum自身に加算して（ステップＳ２４）、変数Cntをインクリメントする（ステップＳ２５）。

【００５１】また、上記ステップＳ２３において、e(f)がnoiselev以下の場合には、そのまま次のステップＳ２６に移る。

【００５２】次に、ファイルが終端に達したか否かを判定し（ステップＳ２６）、まだ終端に達していない場合には上記ステップＳ２２に戻って上述した処理を繰り返す。

【００５３】また、このステップＳ２６においてファイ

ルの終端に達した判断した場合は、上記変数Sumを変数Cntで割ることにより、信号対雑音比の平均値ＳＮ比を計算する（ステップＳ２７）。

【００５４】このように、上記実施形態によれば、よりユーザフレンドリーなディクテーションシステムを提供することができる。

【００５５】[付記]以上詳述した如き本発明の実施形態によれば、以下の如き構成を得ることができる。即ち、

（１）　音声データが記録された記録媒体から音声データを読み出す音声データ読出手段と、上記音声データ読出手段で読み出した音声データの信号対雑音比を演算する信号対雑音比演算手段と、上記信号対雑音比演算手段の出力値が適正であるか否かを判断する信号対雑音比判断手段と、上記信号対雑音比判断手段で上記出力値が適正であると判断した場合に、上記音声データを音声認識処理する音声認識処理手段と、を具備することを特徴とする音声処理装置。

【００５６】（２）　上記（１）に記載の音声処理装置において、上記信号対雑音比判断手段で上記出力値が適正でないと判断した場合に、該適正でない旨を表示する警告手段をさらに具備する。

【００５７】（３）　上記（１）に記載の音声処理装置において、上記信号対雑音比判断手段で上記出力値が適正でないと判断した場合に、上記音声データに音声認識処理を施すか否かを選択する選択手段をさらに具備する。

【００５８】（４）　上記（１）に記載の音声処理装置において、上記信号対雑音比判断手段で上記出力値が適正でないと判断した場合に、該適正でない旨を表示する警告手段と、上記音声データに音声認識処理を施すか否かを選択する選択手段と、をさらに具備する。

【００５９】（５）　上記（１）～（４）に記載の音声処理装置において、上記音声認識処理手段の認識結果を出力する出力手段を具備する。

【００６０】この出力手段としては、ディスプレイ５の他、プリンタ等が相当する。

【００６１】（６）　上記（１）～（５）に記載の音声処理装置において、上記信号対雑音比判断手段は、上記信号対雑音比演算手段の出力値と初期設定基準値とを比較する比較手段を具備する。

【００６２】（７）　コンピュータによって音声認識プログラムに対して音声データを渡す処理をするための処理プログラムを記録した記録媒体であって、上記処理プログラムは、コンピュータに上記音声データが記録された記録媒体より上記音声データを読み込ませ、読み出した上記音声データの信号対雑音比を演算させ、信号対雑音比の値が適正かどうかを判断させ、信号対雑音比の値が適正である場合に、上記音声データを音声認識プログラムに対して渡させることを特徴とする処理プログラムを記録した記録媒体。

【００６３】（８）　コンピュータによって音声認識プログラムに対して音声データを渡す処理をするための処理プログラムを記録した記録媒体であって、上記処理プログラムは、コンピュータに上記音声データが記録された記録媒体より上記音声データを読み込ませ、読み出した上記音声データの信号対雑音比を演算させ、信号対雑音比の値が適正かどうかを判断させ、信号対雑音比の値が適正でない場合に、操作者に上記音声データの音声認識処理を行うかどうかを選択させ、信号対雑音比の値が適正である場合及び操作者が音声認識処理の実行を選択した場合に、上記音声データを音声認識プログラムに対して渡させることを特徴とする処理プログラムを記録した記録媒体。

【００６４】（９）　コンピュータによって音声認識をするための音声認識プログラムを記録した記録媒体であって、上記プログラムは、コンピュータに上記音声データが記録された記録媒体より上記音声データを読み込ませ、読み出した上記音声データの信号対雑音比を演算させ、信号対雑音比の値が適正かどうかを判断させ、信号対雑音比の値が適正でない場合に、操作者に上記音声データの音声認識処理を行うかどうかを選択させ、信号対雑音比の値が適正である場合及び操作者が音声認識処理の実行を選択した場合に、上記音声データを音声認識させることを特徴とする音声認識プログラムを記録した記録媒体。

【００６５】

【発明の効果】以上説明したように本発明によれば、よりユーザフレンドリーな音声処理装置を提供できる。

【図面の簡単な説明】

【図１】本発明の一実施形態のディクテーションシステムの概念的な全体構成を示した図である。

【図２】上記実施形態のディクテーションシステムにおけるパーソナルコンピュータの電気的な構成を示すブロック図である。

【図３】上記実施形態のディクテーションシステムにおいて、音声メモリから音声データを読み出して音声認識するときの全体の流れを示す概念図である。

【図４】上記実施形態のディクテーションシステムにおける音声認識処理を示すフローチャートである。

【図５】上記実施形態のディクテーションシステムにおける音声データの雑音レベルの計算手法を示したフローチャートである。

【図６】上記実施形態のディクテーションシステムにおけるＳＮ比を計算する処理の内容を示したフローチャートである。

【符号の説明】

１…ディジタルレコーダ
２…ミニチュアカード（記録媒体）
４…パーソナルコンピュータ（音声処理装置）
５…ディスプレイ（出力手段、警告手段）

9

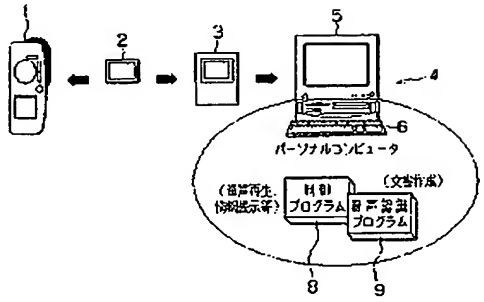
6…キーボード（選択手段）
7…マウス（選択手段）
8…制御プログラム
9…音声認識プログラム
１１…音声メモリ（記録媒体）
１２…復号化処理
１３…ＳＮ比計算処理
１４…判定処理

10

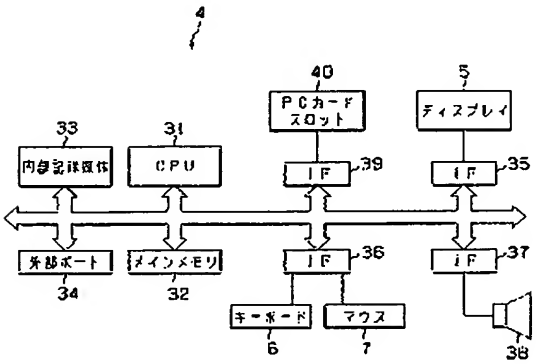
＊１５…音声認識処理
１６…表示
３１…ＣＰＵ（音声データ読出手段．信号対雑音比演算手段．信号対雑音比判断手段、音声認識処理手段）
３２…メインメモリ（記録媒体）
３３…内部記録媒体（記録媒体）
３８…スピーカ（警告手段）

＊

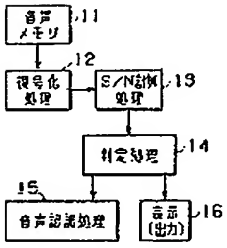
【図１】



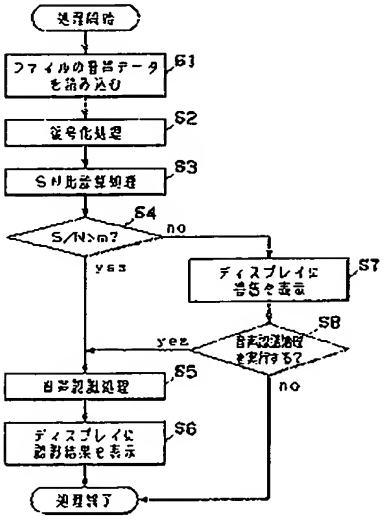
【図２】



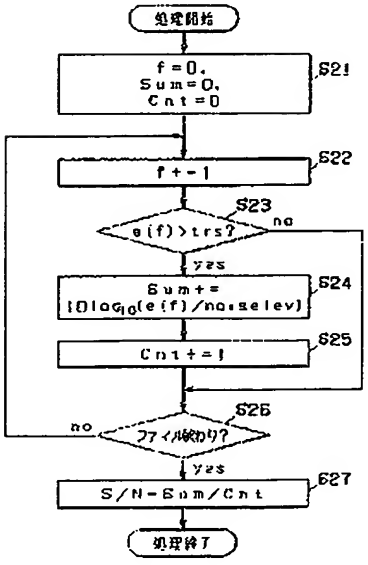
【図３】



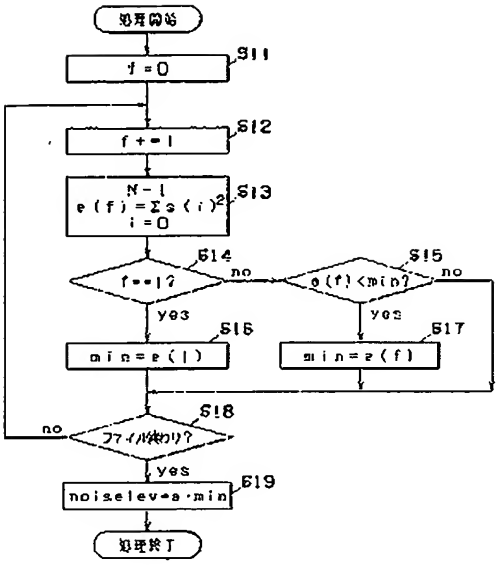
【図４】



【図６】



【図５】



フロントページの続き

(51)Int.Cl.'	識別記号	Ｆｉ		ﾃｰﾏｺｰﾄﾞ(参考)
Ｇ１０Ｌ　19/00		Ｇ１０Ｌ　9/18	Ｊ	

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-056797
(43)Date of publication of application : 25.02.2000

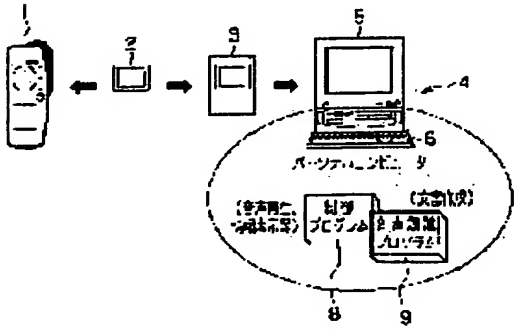
(51)Int.Cl. G10L 15/28
G10L 13/00
G10L 11/02
G06F 3/16
G10L 11/00
G10L 19/00

(21)Application number : 10-224927
(22)Date of filing : 07.08.1998

(71)Applicant : OLYMPUS OPTICAL CO LTD
(72)Inventor : TAKAHASHI HIDEYUKI

(54) SPEECH PROCESSING DEVICE

(57)Abstract:
PROBLEM TO BE SOLVED: To provide a more user-friendly speech processing device.
SOLUTION: This speech processing device reads a speech data, which has been recorded in a miniature card 2 and recorded by a digital recorder 1, by a personal computer 4; calculates SN ratio of the speech data read by this personal computer 4 and judges whether or not the value of the SN ratio is proper; and if it is proper, the speech processing device recognizes the speech data by a speech recognition program 9.



LEGAL STATUS
[Date of request for examination]
[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

* NOTICES *

CLAIMS

[Claim(s)]

[Claim 1] The speech processing unit carry out providing the speech-recognition processing means which carries out the speech-recognition processing of the above-mentioned voice data when it is judged from the record medium with which voice data was recorded with the voice-data read-out means which reads voice data, a signal-to-noise-ratio operation means calculate the signal-to-noise ratio of the voice data read with the above-mentioned voice-data read-out means, a signal-to-noise-ratio decision means judge that the output value of the above-mentioned signal-to-noise-ratio operation means is proper, and the above-mentioned signal-to-noise-ratio decision means that the above-mentioned output value is proper as the description.

[Claim 2] the case where it is judged with the above-mentioned signal-to-noise-ratio decision means that the above-mentioned output value is

not proper -- this -- the speech processing unit according to claim 1 characterized by providing further a warning means to display the purport which is not proper.
[Claim 3] The speech processing unit according to claim 1 characterized by providing further a selection means to choose as the above-mentioned voice data whether speech recognition processing is performed when it is judged with the above-mentioned S/N decision means that the above-mentioned output value is not proper.

[Translation done.]

* NOTICES *

JPO and INPIT are not responsible for any
damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

DETAILED DESCRIPTION

[Detailed Description of the Invention]
[0001]
[Field of the Invention] This invention relates to a speech processing unit and the speech processing unit which performs predetermined processing to the voice data read from the record medium in detail.
[0002]
[Description of the Prior Art] The so-called dictate system implementation which will draw up a document automatically based on this voice data, and will display it on a screen etc. if voice data is inputted the so-called voice word processor or by stating orally is one target in the speech recognition system development from the former, and research is advanced actively now.
[0003] The equipment which connects a microphone to a personal computer, documents the voice inputted using this microphone on this personal computer with an advance of such a speech recognition technique in recent years and a computer technique, and is displayed on a screen is developed, and, generally it is marketed.

[0004] Oral statement sound recording of the content of the document [in drawing up a document] to draw up conventionally on the other hand was once carried out at sound recording equipments, such as a tape recorder, and it has become common as one of the effective utilization gestalten of sound recording equipments, such as a tape recorder, to take the gestalt of documenting with document preparation equipments, such as a typewriter and a word processor, while a secretary, a typist, etc. reproduce the content of oral statement later.
[0005] Implementation of the technique of changing the content of sound recording into a document from before automatically in such oral statement sound recording is desired strongly.
[0006] Moreover, it is possible for the so-called digital recorder which digital-data-izes the content of sound recording, and is recorded on the record medium in which the writing of a flash memory etc. and elimination are possible by development of computer technology in recent years, a digital-signal-processing technique, etc. to comes to be developed, to transmit the digitized content of sound recording to a personal computer further, and to reproduce the content of sound recording in this personal computer.
[0007] These people are developing the processing control unit of the voice data which makes it possible to set simply the sound recording data transmitted from such a digital recorder on a personal computer, and to treat them by easy actuation, and have proposed in Japanese Patent Application No. No. 149728 [nine to].
[0008] Furthermore, these people pass and do speech recognition of the voice by which digital recording was carried out to a voice recognition unit from the processing control device of the above-mentioned voice data, the dictate system displayed on a screen as a document is developed, and it has proposed in Japanese-Patent-Application-No. 9-149729 No.
[0009] According to such a dictate system, it is not necessary to sit down in front of a computer and to carry out a direct sound voice input, it once records into a digital recorder, and it becomes possible to transmit the sound recording data to a computer later, and to make a document draw up.
[0010]
[Problem(s) to be Solved by the Invention] By the way, the large lexical continuous speech recognition technique for unspecified speakers is needed for speech recognition processing in a dictate system which was mentioned above. However, in the current large lexical continuous

speech recognition technique for unspecified speakers, obtaining a perfect recognition result without incorrect recognition has the problem that the recognition engine performance deteriorates, when it is very difficult and a background noise mixes into the voice for recognition especially. In order to solve such a problem conventionally, it is just going to be known well that various proposals are made. However, it is difficult to solve it with limited equipment.

[0011] If it is when using the equipment which connects a microphone to a personal computer which was mentioned above, documents the voice inputted using this microphone on this personal computer, and is displayed on a screen in such the actual condition, the speech-recognition result displayed on a screen on that spot checks, and it is decision of a user, and if there is much incorrect recognition, it is possible to also take response of redoing voice input again.

[0012] In the dictate system displayed on a screen on the other hand by using as a document the result of having made carrying out voice data which was mentioned above, and by which digital recording was carried out to a processing control unit to a voice recognition unit, and having recognized speech recognition, the already recorded voice data serves as an input to a voice recognition unit.

[0013] Therefore, when a loud background noise performs speech recognition processing to the voice data mixed and recorded, a recognition result with it there is much incorrect recognition and difficult [to even correct later] will be displayed, and even if it reperforms speech recognition processing again by decision of a user, there is a problem that there is no chance that a recognition result will be improved.

[0014] The place which a dictate system which was mentioned above makes the original object is for it to be more quick and document the content of the recorded voice data more simply, i.e., offer document preparation exchange. Since what is necessary is just to correct the incorrect recognition part using a keyboard, a mouse, etc. if incorrect recognition parts are few as a result of the speech recognition processing to the recorded voice data, the object can be attained.

[0015] However, if an incorrect recognition part increases above to some extent, even correction will become difficult and will be referred to as being able to draw up a document having retyped quickly [way] from the start after all. A user will be burdened with big disadvantage in saying [that it does not turn out that it is such unless it processes actually and achieves results].

[0016] this invention is made in view of this trouble -- having -- more -- user FUREN -- it aims at offering a dolly speech processing unit.
[0017]
[Means for Solving the Problem] In order to attain the above-mentioned object the 1st speech processing unit of this invention The voice data read-out means which reads voice data from the record medium with which voice data was recorded, A signal-to-noise-ratio operation means to calculate the signal-to-noise ratio of the voice data read with the above-mentioned voice data read-out means, A signal-to-noise-ratio decision means to judge whether the output value of the above-mentioned signal-to-noise-ratio operation means is proper, and the speech recognition processing means which carries out speech recognition processing of the above-mentioned voice data when it is judged with the above-mentioned signal-to-noise-ratio decision means that the above-mentioned output value is proper are provided.
[0018] the case where the 2nd speech processing unit of this invention judges that the above-mentioned output value is not proper with the above-mentioned signal-to-noise-ratio decision means in the 1st speech processing unit of the above in order to attain the above-mentioned object -- this -- a warning means to display the purport which is not proper is provided further.
[0019] In order to attain the above-mentioned object, the 3rd speech processing unit of this invention possesses further a selection means to choose as the above-mentioned voice data whether speech recognition processing is performed, in the 1st speech processing unit of the above, when it is judged with the above-mentioned S/N decision means that the above-mentioned output value is not proper.
[0020]
[Embodiment of the Invention] Hereafter, the gestalt of operation of this invention is explained with reference to a drawing.
[0021] Drawing 1 thru/or drawing 6 start the dictate system which is 1 operation gestalt of this invention, and drawing 1 is drawing having shown the notional whole configuration of this dictate system.
[0022] The digital recorder 1 which changes voice into an electrical signal and voice-data-izes it as this dictate system is shown in drawing 1 , Record-medium slack Miniature Card 2 which equips this digital recorder 1 removable, is used for it, and records the above-mentioned voice data, The PC card adapter 3 for inserting in PC Card slot 40 (referring to drawing 2) which mentions this Miniature Card 2 later, and making connection possible, To the voice data which was equipped with

the output means slack display 5, the keyboard 6, and the mouse 7 grade, and was obtained from above-mentioned Miniature Card 2 through above-mentioned PC Card slot 40 It has the personal computer 4 as a speech processing unit which performs processing by the control program 8 or the speech recognition program 9, and is constituted.
[0023] In addition, the above-mentioned display 5 achieves the duty as a warning means to display the purport which is not proper.
[0024] Drawing 2 is the block diagram showing the electric configuration of the above-mentioned personal computer 4.
[0025] While this personal computer 4 performs voice playback, an information display, etc. according to the above-mentioned control program 8 and performs document preparation etc. according to the above-mentioned speech recognition program 9 While performing various processings according to various kinds of other programs and managing control of the personal computer 4 whole concerned CPU31 which achieves duties, such as a voice data read-out means, a signal-to-noise-ratio operation means, a signal-to-noise-ratio decision means, and a speech recognition processing means, The record-medium slack main memory 32 used as the working area of this CPU31, For example, the interior record medium 33 of record-medium slack with which it becomes by the hard disk, a floppy disk, etc., and the above-mentioned control program 8 and the speech recognition program 9 are recorded, The external port 34 for connecting with various kinds of external instruments, and the interface 35 which connects the above-mentioned display 5 (it abbreviates to IF hereafter), IF36 which connects the above-mentioned keyboard 6 and a mouse 7, and the loudspeaker 38 which utters voice based on voice data, It has IF37 which connects this loudspeaker 38, PC Card slot 40 in which Miniature Card 2 with which the above-mentioned PC card adapter 3 was equipped is inserted, and IF39 for connecting this PC Card slot 40, and is constituted.
[0026] Moreover, the above CPU 31, main memory 32, the internal record medium 33, the external port 34, and IF 35, 36, 37, and 39 are mutually connected through the bus.
[0027] In addition, although you may make it read from Miniature Card 2 directly through above-mentioned PC Card slot 40, voice data is recorded on the above-mentioned internal record medium 33, and it may be made to once read it from this internal record medium 33, or even if it makes it read from the digital recorder 1 directly through means of communications etc., it is not cared about.
[0028] Moreover, the above-mentioned loudspeaker 38 achieves the

duty as a warning means which displays the purport which is not proper (pronunciation).

[0029] Next, the speech recognition processing in the dictate system of this operation gestalt is explained with reference to drawing 3 and drawing 4 .

[0030] Drawing 3 is the conceptual diagram showing the whole flow when reading and carrying out speech recognition of the voice data from voice memory in the dictate system of this operation gestalt, and drawing 4 is a flow chart which shows processing of the speech recognition in this dictate system.

[0031] If speech processing is started as shown in drawing 4 , the voice data currently recorded in the file unit will be read from the voice memory 11 (record medium with which voice data was recorded) as a record medium with which above-mentioned Miniature Card 2 or the above-mentioned internal record-medium 33 grade, and voice data were recorded (step S1), and decryption processing 12 will be performed (step S2).

[0032] The result of this decryption processing 12 calculates an SN ratio by the technique which it is sent to the SN ratio (signal-to-noise ratio) computation 13, for example, is mentioned later (step S3).

[0033] The calculated value (S/N) of this SN ratio is compared with the predetermined value m by the judgment processing 14 (step S4). In this step S4, if it is S/N>m, voice data will be sent to the speech recognition processing 15, and speech recognition will be performed (step S5). And the result of this speech recognition is displayed on the screen of display 5 grade (step S6).

[0034] By the above-mentioned step S4, if it is not S/N>m, warning of the purport that the recognition result which may continue speech recognition as it is will not be obtained will be displayed on a display 5 (step S7).

[0035] In addition, this alarm display may not be restricted for making it display on a display 5, for example, you may warn of it with voice etc. from a loudspeaker 38.

[0036] After the above-mentioned warning, speech recognition processing is performed as it is, activation of processing is stopped, or selection is demanded from a user (step S8). In addition, also by the key stroke on a keyboard 6, this selection is good and may be operated by mouse 7 grade.

[0037] As a result of selection by the above-mentioned user, if it is yes, it will carry out to step S5 and speech recognition processing will be

performed. On the other hand, processing will be ended if it is no as a result of selection by the user.
[0038] Next, the content of processing of the SN ratio computation in the above-mentioned step S3 is explained with reference to the flow chart shown in drawing 5 and drawing 6 .
[0039] First, it explains with reference to the flow chart which shows the count technique of the noise level of voice data to drawing 5 .
[0040] If this processing starts, the variable f which shows the counted value of a frame number will be first initialized to 0 (step S11).
[0041] Next, after incrementing Variable f, frame energy e (f) is calculated with (step S12) and the formula of a graphic display (step S13). In addition, an input signal [in / in s (i) / the sample of eye (i) watch in one frame] and N show among the formula the measurement size which constitutes one frame.
[0042] Next, it judges whether it is the frame of whether the value of Variable f is 1, and the first stage (step S14), and when f is 1, the value of the variable min which shows the minimum frame energy is set to e (1) (step S16).
[0043] Moreover, when f is not 1 in the above-mentioned step S14, it judges whether frame energy e (f) is smaller than Variable min (step S15), in being small, it sets frame energy e (f) to Variable min (step S17), and in not being small, on the other hand, it goes to the following step S18, without doing anything as it is.
[0044] And it judges whether the file reached termination (step S18), and in not being termination still, it repeats the processing which returned and mentioned above to the above-mentioned step S12.
[0045] Moreover, when it is judged that the end of file was reached in this step S18, the value which integrated the predetermined value alpha (for example, 1.8) is set to the above-mentioned variable min as threshold noiselev (step S19), and it escapes from this processing.
[0046] Such a detection approach of noise level uses effectively that voice data is already recorded, and becomes possible [**(ing) to count of an accurate SN ratio].
[0047] In addition, although the minimum value of the read entire interval (that is, all frames that constitute a voice file) is calculated in ****, even if this invention is not limited to this and is no minimum value of the sections, it should just be the section of a certain amount of die length.
[0048] Next, it explains with reference to the flow chart which shows the content of the processing which calculates an SN ratio to drawing 6 .

[0049] If this processing starts, the variable Cnt which shows the variable Sum which shows the aggregate value of the variable f which shows the counted value of a frame number, and the signal-to-noise ratio (SN ratio) of each frame, and the count of addition will be respectively initialized to "0" (step S21).
[0050] Next, frame energy e (f) which calculated Variable f in drawing 5 incremented and (step S22) mentioned above judges whether it is larger than noise level noiselev (step S23). In being larger than noiselev, e (f) adds the calculated value of the signal-to-noise ratio of the frame concerned to the variable Sum itself (step S24), and increments Variable Cnt here (step S25).
[0051] Moreover, in the above-mentioned step S23, when e (f) is below noiselev, it moves to the following step S26 as it is.
[0052] Next, the processing which returned and mentioned above whether the file reached termination to the above-mentioned step S22 when it judged (step S26) and termination was not reached yet is repeated.
[0053] Moreover, when [which reached the end of file in this step S26] it judges, the average SN ratio of a signal-to-noise ratio is calculated by dividing the above-mentioned variable Sum by Variable Cnt (step S27).
[0054] thus -- according to the above-mentioned operation gestalt -- more -- user FUREN -- a dolly dictate system can be offered.
[0055] According to the operation gestalt of **** this invention explained in full detail more than the [additional remark], the configuration like a less or equal can be obtained. Namely, (1) The voice data read-out means which reads voice data from the record medium with which voice data was recorded, A signal-to-noise-ratio operation means to calculate the signal-to-noise ratio of the voice data read with the above-mentioned voice data read-out means, A signal-to-noise-ratio decision means to judge whether the output value of the above-mentioned signal-to-noise-ratio operation means is proper, The speech processing unit characterized by providing the speech recognition processing means which carries out speech recognition processing of the above-mentioned voice data when it is judged with the above-mentioned S/N decision means that the above-mentioned output value is proper.
[0056] (2) the case where it is judged with the above-mentioned signal-to-noise-ratio decision means in the speech processing unit of a publication to the above (1) that the above-mentioned output value is not proper -- this -- provide further a warning means to display the

purport which is not proper.
[0057] (3) When it is judged with the above-mentioned S/N decision means in a speech processing unit given in the above (1) that the above-mentioned output value is not proper, provide further a selection means to choose as the above-mentioned voice data whether speech recognition processing is performed.
[0058] (4) the case where it is judged with the above-mentioned signal-to-noise-ratio decision means in the speech processing unit of a publication to the above (1) that the above-mentioned output value is not proper -- this -- provide further a warning means to display the purport which is not proper, and a selection means to choose as the above-mentioned voice data whether speech recognition processing is performed.
[0059] (5) The above (1) An output means to output the recognition result of the above-mentioned speech recognition processing means to - (4) in the speech processing unit of a publication is provided.
[0060] A printer besides a display 5 etc. corresponds as this output means.
[0061] (6) The above (1) In a speech processing unit given in - (5), the above-mentioned signal-to-noise-ratio decision means possesses a comparison means to compare the output value and initialization reference value of the above-mentioned signal-to-noise-ratio operation means.
[0062] It is the record medium which recorded the processing program for carrying out processing which passes voice data to a speech recognition program by computer. (7) The above-mentioned processing program The above-mentioned voice data is made to read from the record medium with which the above-mentioned voice data was recorded on the computer. The record medium which is made to calculate the signal-to-noise ratio of the read above-mentioned voice data, was made to judge whether the value of a signal-to-noise ratio is proper, and recorded the processing program characterized by making the above-mentioned voice data pass to a speech recognition program when the value of a signal-to-noise ratio was proper.
[0063] It is the record medium which recorded the processing program for carrying out processing which passes voice data to a speech recognition program by computer. (8) The above-mentioned processing program The above-mentioned voice data is made to read from the record medium with which the above-mentioned voice data was recorded on the computer. Make the signal-to-noise ratio of the read

above-mentioned voice data calculate, and it is made to judge whether the value of a signal-to-noise ratio is proper. When the value of the S/N is not proper, it is made to choose whether speech recognition processing of the above-mentioned voice data is given to an operator. The record medium which recorded the processing program characterized by making the above-mentioned voice data pass to a speech recognition program when the value of a signal-to-noise ratio is proper, and when activation of speech recognition processing of an operator is chosen.

[0064] It is the record medium which recorded the speech recognition program for carrying out speech recognition by computer. (9) The above-mentioned program The above-mentioned voice data is made to read from the record medium with which the above-mentioned voice data was recorded on the computer. Make the signal-to-noise ratio of the read above-mentioned voice data calculate, and it is made to judge whether the value of a signal-to-noise ratio is proper. When the value of the S/N is not proper, it is made to choose whether speech recognition processing of the above-mentioned voice data is given to an operator. The record medium which recorded the speech recognition program characterized by carrying out speech recognition of the above-mentioned voice data when the value of a signal-to-noise ratio is proper, and when activation of speech recognition processing of an operator is chosen.

[0065]

[Effect of the Invention] according to [as explained above] this invention -- more -- user FUREN -- a dolly speech processing unit can be offered.

[Translation done.]

* NOTICES *

JPO and INPIT are not responsible for any
damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]
[Drawing 1] It is drawing having shown the notional whole configuration of the dictate system of 1 operation gestalt of this invention.
[Drawing 2] It is the block diagram showing the electric configuration of the personal computer in the dictate system of the above-mentioned operation gestalt.
[Drawing 3] In the dictate system of the above-mentioned operation gestalt, it is the conceptual diagram showing the whole flow when reading and carrying out speech recognition of the voice data from voice memory.
[Drawing 4] It is the flow chart which shows the speech recognition processing in the dictate system of the above-mentioned operation gestalt.
[Drawing 5] It is the flow chart which showed the count technique of the noise level of the voice data in the dictate system of the above-mentioned operation gestalt.
[Drawing 6] It is the flow chart which showed the content of the processing which calculates the SN ratio in the dictate system of the above-mentioned operation gestalt.
[Description of Notations]
1 -- Digital recorder
2 -- Miniature Card (record medium)
4 -- Personal computer (speech processing unit)
5 -- Display (an output means, warning means)
6 -- Keyboard (selection means)
7 -- Mouse (selection means)
8 -- Control program
9 -- Speech recognition program
11 -- Voice memory (record medium)
12 -- Decryption processing
13 -- SN ratio computation
14 -- Judgment processing
15 -- Speech recognition processing
16 -- Display
31 -- CPU (a voice data read-out means, a signal-to-noise-ratio operation means, a signal-to-noise-ratio decision means, speech recognition processing means)
32 -- Main memory (record medium)
33 -- Internal record medium (record medium)

38 -- Loudspeaker (warning means)

[Translation done.]

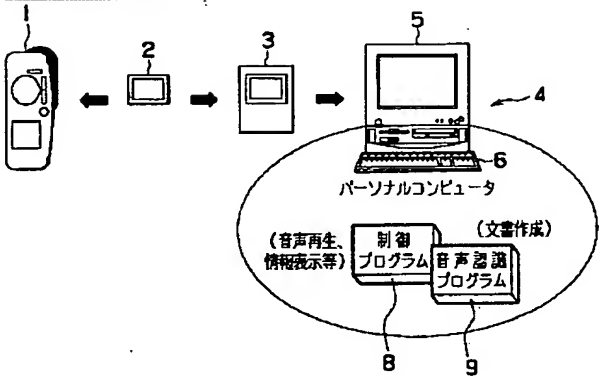
* NOTICES *

JPO and INPIT are not responsible for any
damages caused by the use of this translation.

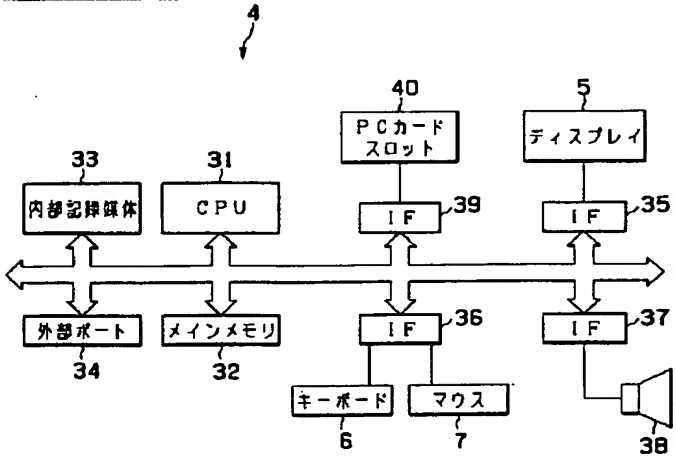
1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

DRAWINGS

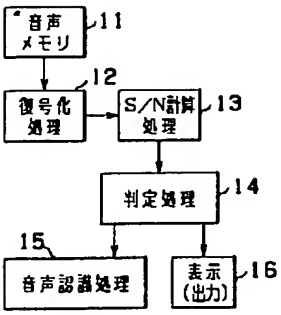
[Drawing 1]



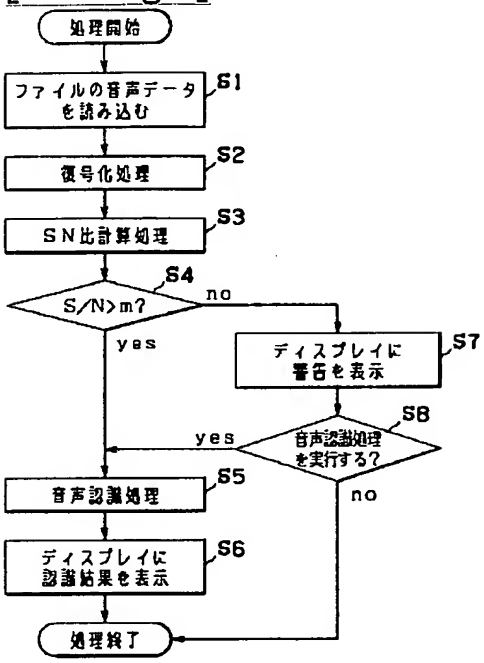
[Drawing 2]



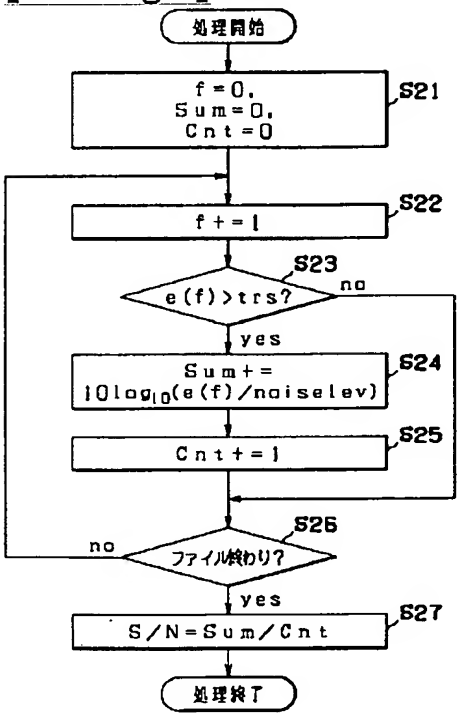
[Drawing 3]



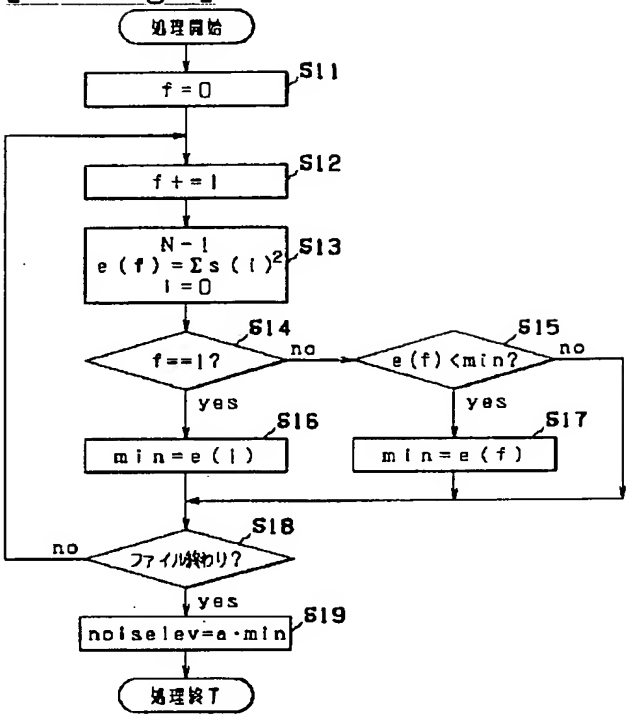
[Drawing 4]



[Drawing 6]



[Drawing 5]



[Translation done.]